CONSTRUCTION OF SYMMETRY TRIANGULAR FUZZY NUMBER PROCEDURE (STFNP) USING STATISTICAL INFORMATION FOR AUTOREGRESSIVE FORECASTING

MUHAMMAD SHUKRI BIN CHE LAH

A thesis submitted in
fulfillment of the requirement for the award of the
Master of Information Technology

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

JANUARY 2020

For my beloved father Che Lah Bin Abu and mother Halijah Binti Ahmad.

ACKNOWLEDGEMENT

In the Name of ALLAH, the Most Gracious and the Most Merciful

The author would like to express his utmost gratitude to all of the people who have supported him throughout this research. He is greatly indebted to his supervisor, Dr. Nureize Binti Arbaiy and co-supervisor, Dr. Riswan Effendi who have helped, in gaving inspirational suggestions and motivation to perform the best of his ability.

This thankfulness is also dedicated to his parents, Che Lah Bin Abu and Halijah Binti Ahmad, his brother Saiful Azrul, Mohamad Shaukhi, Mohamad Faizol, Mohamad Fazli and Muhammad Farizzudin for their tender love, encouragement, and support that have given him so much strength to complete this research successfully.

Last but not least, a very special gratitude to his best friends Thariq, Hafizie, Aslam, Hafizul, Haikal, Nazim, Shuhaidi and Norida who have gone through ups and downs together with him until the very last moment of completing this research.

ABSTRACT

Single-point data are used for data collection. However, data collected by various data collection methods are often exposed to uncertainties that may affect the information presented by the quantitative results. This also causes the forecast model developed to be less precise because of the uncertainties contained in the input data. It is essential to describe the uncertainty in data to obtain a realistic result from data analysis. However, most studies focus on model uncertainty regardless of data uncertainty. The data processing carried out may not always take care of uncertainty. When uncertainties in the raw data are not sufficiently handled, this creates more errors that are included in the predicted model. Standard procedures are also very limited to be followed in order to transform a single-point value into Triangular Fuzzy Number (TFN), which addresses the uncertainty. Thus, the data preparation procedure of Symmetry Triangular Fuzzy Number (STFN) is presented in this study to build an improved autoregressive model for time series forecasting. This study presents the proposed Symmetry Triangular Fuzzy Number Procedure (STFNP) using percentage error method and standard deviation method for first-order autoregressive forecasting. Percentage error rate method involves three different percentage rates, while the second method uses the standard deviation of the data. Simulations and verification procedures are presented and are accompanied with numerical examples using actual datasets of Air Pollutant Index and stock markets of selected ASEAN countries. This study reveals that the percentage error and standard deviation methods, which were used to construct the TFN, can achieve the same or better accuracy as compared to a single-point procedure. The results of the simulations and experiments show that the standard deviation method produces better results compared to the other proposed approaches and the conventional approach. Besides, the systematic procedure to construct the TFN does not deviate from single-point procedures. Importantly, uncertain data being treated avoids more uncertainties that would have been brought to the outcome of the forecast model and consequently improves prediction accuracy.

ABSTRAK

Data titik tunggal digunakan bagi tujuan pengumpulan data. Walau bagaimanapun, data yang dikumpul melalui pelbagai kaedah pengumpulan data sering terdedah kepada ketidakpastian yang mungkin mempengaruhi maklumat yang dibentangkan oleh hasil kuantitatif. Hal ini juga mengakibatkan model ramalan yang dibangunkan menjadi kurang tepat kerana ketidakpastian yang terkandung dalam data input. Adalah penting bagi kita untuk mengolah ketidakpastian dalam data bagi mendapatkan hasil yang realistik daripada analisis data. Namun, kebanyakan kajian yang menumpukan pada ketidakpastian model tidak mengambil kira permasalahan ketidakpastian data. Sesuatu kaedah pemprosesan data yang dijalankan mungkin tidak mengambil kira perihal ketidakpastian data. Apabila ketidakpastian dalam data mentah tidak dikendalikan secukupnya, hal ini mencetus lebih banyak kesilapan dimasukkan ke dalam model ramalan. Prosedur standard bagi menangani ketidakpastian juga adalah sangat terhad untuk diikuti bagi mengubah nilai titik tungal ke dalam nilai *Triangular Fuzzy Number* (*TFN*). Oleh itu, prosedur penyediaan data bagi *Symmetry Triangular Fuzzy Number* (*STFN*) telah dibentangkan dalam kajian ini bagi membina model *autoregressive* yang dipertingkatkan bagi tujuan ramalan siri masa. Kajian ini membentangkan cadangan bagi *Symmetry Triangular Fuzzy Number Procedure* (*STFNP*) menggunakan kaedah ralat peratusan dan kaedah sisihan piawai bagi peramalan *autoregressive* tahap pertama. Kaedah ralat peratusan melibatkan tiga kadar peratusan yang berbeza, manakala kaedah kedua menggunakan nilai sisihan piawaian data. Prosedur simulasi dan pengesahan dibentangkan dan diiringi dengan contoh numerikal menggunakan set data sebenar bagi Indeks Pencemaran Udara dan pasaran saham bagi negara-negara ASEAN yang terpilih. Kajian ini mendedahkan bahawa peratusan kadar kesilapan dan kaedah sisihan piawai yang digunakan untuk membina TFN mampu mencapai tahap ketepatan yang sama atau lebih baik berbanding prosedur titik tunggal. Hasil simulasi dan eksperimen menunjukkan bahawa kaedah sisihan piawai menghasilkan keputusan yang lebih baik berbanding kaedah lain yang

dicadangkan termasuk kaedah konvensional. Selain itu, prosedur sistematik untuk membina TFN tidak menyimpang dari prosedur titik tunggal. Yang penting, ketidakpastian data yang telah dirawat mengelakkan lebih banyak ketidakpastian dibawa kepada model ramalan dan seterusnya meningkatkan ketepatan ramalan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

e_t	-	White noise (error)
$\emptyset_p$	-	Coefficient
y_{t-p}	-	Past series time series
y_t	-	Time series at time, t
c	-	Constant
C	-	Center value of TFN
α	-	Spread of TFN
$\tilde{y}$	-	Non-fuzzy number
α_r	-	Right spread
α_l	-	Left spread
$\hat{X}_i$	-	Observed value
X_i	-	Predicted values
n	-	Number of sample
$\overleftarrow{\boldsymbol{y}}_{\boldsymbol{t}}^{\Delta}$	-	Left predicted value
$\overrightarrow{\boldsymbol{y}}_{\boldsymbol{t}}^{\Delta}$	-	Right predicted value
$\widetilde{\overrightarrow{\boldsymbol{y}}}_{\boldsymbol{t}}^{\Delta}$	-	Centre point of STFN
$\tilde{y}_t^p$	-	Fuzzy time series data at time, t
p	-	Spread for percentage error
S	-	Spread for standard deviation
σ	-	Standard deviation
$\bar{x}$	-	Mean of x_i
MSE	-	Mean Squared Error
TFN	-	Triangular Fuzzy Number
STFN	-	Symmetry Triangular Fuzzy Number
STFNP	-	Symmetry Triangular Fuzzy Number Procedure
FN	-	Fuzzy Number
FLE	-	First Level Error
SLE	-	Second Level Error
AR	-	Autoregressive
AR(1)	-	First Order Autoregressive
SD	-	Standard Deviation
PE	-	Percentage Error
API	-	Air Pollutant Index
ASEAN	-	Association of Southeast Asian Nations
OLS	-	Ordinary Least Squares
CSV	-	comma-separated values
Argen	-	Autoregressive Generator

AC	-	Ant Colony
FFA	-	Firefly Algorithm
CSA	-	Cuckoo Search, Algorithm
GA	-	Genetic Algorithm
ARIMA	-	Autoregressive integrated moving average
MA	-	Moving average
ARMA	-	Autoregressive moving average

LIST OF APPENDICES

LIST OF PUBLICATIONS

Journals:

(i) Muhammad Shukri Che Lah, Mohammad Haris Haikal Othman, Nureize Arbaiy, First Order of Autoregressive Air Pollution Forecasting with Symmetry Triangular Fuzzy Number based on Percentage Error, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Vol. 8, Issue-8S, pp. 265 – 269, 2019.

(ii) Muhammad Shukri Che Lah, Nureize Arbaiy, Triangular Fuzzy Number Generator (TriGen), International Journal of Advanced Trends in Computer Science and Engineering, ISSN: 2278-3091, Vol. 8, Issue-1.3, pp. 300 – 365, 2019.

Proceedings:

(i) Lah, M. S. C., Arbaiy, N., & Efendi, R. (2019). Stock Market Forecasting Model Based on AR (1) with Adjusted Triangular Fuzzy Number Using Standard Deviation Approach for ASEAN Countries. In *Intelligent and Interactive Computing*(pp. 103-114). Springer, Singapore.

LIST OF AWARDS

CHAPTER 1

INTRODUCTION

1.1 Thesis Background

The knowledge of the physical world is understood and theoretically proven by simulating the real-world environment using a closed-condition experiment with real-world data. Some experiments may involve activities such as measurements, test methods, and experimental designs, which are used to gather the data. There are two types of data collected, which are primary data and secondary data [1], [2].

Real-world data are derived from a variety of sources [3]–[5], which are associated to represent the whole sample space. In order to extract meaningful information that lies within the data, the first approach is to analyse the information. Data analysis involves the process of data checking, data cleaning, data transformation, and data modelling for research and decision-making purposes [6]. Data analysis assists the process of determining any patterns, relationship, or trends and turns them into useful information which finally is useful in decision making [7], [8]. Understanding the derived data from experiments and the real world generates meaningful insights into the various needs for the decision-making process.

Unfortunately, most of the existing studies on modelling [9]–[15] are focused on building a model without giving attention to the reliability of input data—from either primary sources or secondary sources. For instance, the response variable (input data) in the previous period has become the predictor in a regression model. In data science and analytical process, assembling an appropriate time series data input is considered a crucial task and challenging. In time series forecasting, past observations are collected and analysed to develop a suitable mathematical model which captures the underlying data generating a process for the series [16], [17]. Time series observations are frequently encountered in many domains, such as business,

economics, industry, engineering, and science [16], [17]. There are various types of time series depending on the type of analysis and practical needs.

Before analysis, data preparation procedure is executed first in order to reduce the risk of missing data, imbalanced data, incomplete data, and others. Data preparation is done to avoid misleading results during analysis, which cause forecasting error. Data collected from various measurements carry some degree of uncertainty [18], which may cause a higher risk of data uncertainty.

It is important to describe the uncertainty in data to obtain realistic results from data analysis. One of the most challenging issues is interpreting evidences of observations that contain inherent measurement errors. Observation errors may be derived from the measurement process These errors are the combined measure of the variations that exist in the observed phenomenon and various factors that interfere with the measurements [19], [20]. The inaccurate measured data can potentially lead to a biased estimation [21], [22]. Data with many kinds of uncertainties, such as political uncertainties, risk, incomplete knowledge, and random events, could affect the data's reliability [23]. Such uncertainties may affect the information conveyed by the quantitative result [24]. Although there is an increase in the evidence of the importance of uncertainty, the difficulty in measuring uncertainty remains a challenge in this area [25]. Since it is impossible to eliminate all measurement errors in order to obtain a totally precise value, estimation or approximation measure with certain limits is a considerable approach.

The uncertainties contained in the historical data which is used to build a forecasting model can affect the performance of the model. The existence of inherent uncertainties makes the conventional analysis incapable to deal with such data [26], [27]. For example, time series observation only records single-point data values, which are only best suited to the conventional time series analysis such as autoregressive (AR) models. However, most studies focus on model uncertainty regardless of any data uncertainty. The data processing procedure being carried out may not always handle the uncertainty issue. Given that uncertainties in the input data are not sufficiently handled, this situation may create more errors in the predictive model. Standard procedures for addressing data uncertainty are also very limited to be followed. Since these uncertainties may have substantial implications for interpreting a process, a systematic technique for dealing with uncertainties during data preparation is necessary [28]–[31].

Currently, most researchers often use single-point techniques when modelling various fields [32]–[36]. Single-point data are used to represent multiple data which are observed in sequential order of time, by taking the average of the data. However, single-point data are not considered a variability issue. The average can be minimised by outliers in a way that it does not deal well with varied samples. Apparently, most data are obtained from secondary sources. Since data are derived from secondary sources, legitimacy, bias, and representation may be present and become a challenging issue. This limitation contributes to the construction of a less accurate prediction model.

Less frequently, numerical data could not be determined accurately or precisely because of measurement techniques (human error) or faulty instruments (technical error). Thus, researchers [37], [38] have introduced fuzzy numbers to represent the single-point data in a fuzzy form. By applying fuzzy numbers to represent a single-point data, uncertainty in data and in the developed model are addressed much clearly. However, only a few discuss the fuzzy data treatment during data preparation despite the importance to treat the data before building the forecast model. For example, the decision to identify the spread of the fuzzy number is not explained. In the real study, it is not always the expert who provides information regarding the spread value to form a fuzzy number, hence, causing the fuzzy number's formation to become misleading and the inability to treat the uncertainties sufficiently. On the other hand, the spread of the fuzzy number assumes that only 5% error is allowed. This also leads to an error in data preparation and contributes to less accurate forecasting. Therefore, a new approach in fuzzy data preparation is required to improve the accuracy in forecasting with uncertainties.

In the literature, few research are focused on data preparation. Minimising errors in data collection and preparation may contribute to a better prediction for model building. Measuring the correctness of data collection is necessary to reduce the possibility of hidden errors accompanied in the built model. To address the problems, a new approach of data preparation for time series by adjusting the spread of the symmetry triangular fuzzy number in building autoregressive models is introduced in this research.

1.2 Problem Statement

Data preparation often starts with data collection, where some methods have been used [39]–[41] such as primary data and secondary data. There are several errors which may be involved in data collection, such as systematic error, random error, and data manipulation. Most of the existing studies [10]–[15] on modelling conventional and fuzzy time series analysis have focused on data consisting of numerical values (single point). The sources of data are also not discussed in their research, where there exists uncertainty in the data [42], [43] either from primary data or secondary data.

Figure 1.1 illustrates the current process of AR forecasting. There are some factors that contribute to the uncertainty in data collection, such as systematic error. Systematic error is defined as a component of error which, in the course of several analyses of the same measurements, remains constant or varies in a predictable way [9]. Systematic errors can be compensated if the errors are known. An example is the zero error in a bathroom scale, which causes an incorrect position of the zero point.

Another factor that contributes to the uncertainty is random error, which typically arises from unpredictable variations of influence quantities [9]. It usually results from the experimenter's inability to take the same measurement in the same way to get the same number. Random error can be caused by personal errors, which are human limitation of sight and touch; natural errors, which are changes in temperature while the experiment is in progress; and lack of sensitivity of the instrument, where the instrument fails to respond to the small change [44].

Data manipulation is another factor, where it is the process of changing data to make it easier to read, analyse, or be more organised. In other words, researchers can manipulate data based on specific intentions or goals [45]. This is because data manipulation can cause other irregularities, primarily if other researchers use the same data in their research.

The raw data from this data collection phase are believed to bring First Level Error (FLE), as illustrated in Figure 1.1. When these raw data are used as input data for the forecast, they will cause errors to be brought together in the prediction results.

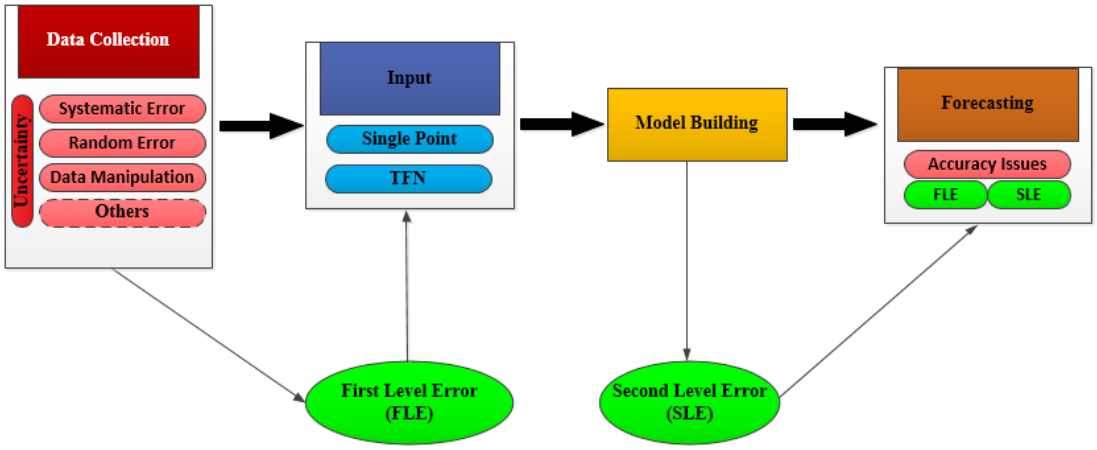

Figure 1.1: Current process of autoregressive (AR) forecasting

In the input phase, the single-point data format is often used by researchers, such as for the average daily observed value of air pollution index. If data are collected many times a day, they may not consider the standard deviation value because the smallest standard deviation means that the data are excellent and reliable. Another example of a single-source data is the closing price or open price from the previous day in the stock market. However, it is difficult to ensure that the average value obtained does not have outliers and, if any, an explanation when calculating the average value, and it indirectly seems to permit an error in the data preparation stage. For instance, in [46], the authors use average data as input, where these data might not represent the nature of the data because the outliers can skew the average.

In fuzzy numbers (FNs), single-point data are known as crisp value. FN has been introduced by Zadeh [47] to deal with uncertainty in a crisp value and imprecise numerical quantities in a practical way. Among the FN methods, TFN is commonly used to convert the crisp value into a fuzzy form, which is known as the Fuzzification process [48]–[50]. Some researchers transform the crisp value into TFN because it gives more options to achieve better accuracy [51]–[54]. Even though the use of TFN has achieved significantly good results, there are few discussions about building TFN [31], [46], [55]. Such discussions which lead to the standard procedure of building an FN are essential as the formation of FN must be built correctly to treat the uncertainty issue [56]. It is critical to have the standard procedure of building FN as it affects the modelling result [57]. Currently, researchers constructing TFN based on expert's assumptions, such as using low-high [27] value or open-close [58] value as the spread of TFN, is inappropriate due to no detailed justification on the selection.

The situation will be worse when the input data, which contain error, are used for forecasting. In this phase, the Second Level Error (SLE) is believed to exist. Eventually, prediction results are affected by errors in both FLE and SLE phases. To overcome these issues, an approach in data preparation is required to improve the accuracy in time series forecasting with uncertainties.

1.3 Objectives of the Study

The objectives of this study are:

(i) To investigate the important parameters in constructing STFNP for time series data.
(ii) To construct a fuzzy autoregressive model using the proposed parameters in (i).
(iii) To evaluate the accuracy of the proposed TFN based fuzzy autoregressive model against the existing TFN.

1.4 Scope of the Study

This study is limited to STFN and first-order autoregressive model to achieve better accuracy for time series analysis. This study provides a systematic procedure for building STFN by introducing percentage error and standard deviation methods. This research is intended to solve the problem of error and uncertainty, which may be contained in the input data, to avoid the error being brought together in the forecasting.

The proposed approaches were validated by the simulation study and actual data applications. The simulation study used generated datasets, which was produced by the Argen simulator tool. Meanwhile, applications on real data have been experimented in two different fields. First, a study was conducted on the data for the Air Pollution Index. The dataset was taken from the Malaysia Air Pollutant Index Management System (APIMS) website, http://apims.doe.gov.my, for the data period ranging from January 1, 2015, to March 30, 2015. Meanwhile, a study involving data from the stock markets used data taken from Investing.com website for the data period from January 1, 2018, to March 26, 2018. Both the simulated experiments and actual applications used a specific model, the AR(1) model, as a fundamental model for autoregression.

1.5 Thesis organization

This report consists of five chapters, including the Introduction and Conclusion chapters. The following is the synopsis of each chapter.

Chapter 1: Introduction.

This chapter describes the research background, problem statements, the objectives, and the scope of this study.

Chapter 2: Literature Review.

This chapter explains the basic concepts of autoregression and FN, as well as applications and multiple work reviews on the methods used by researchers when solving problems related to achieving better accuracy. Besides literature review, critical analysis has also been done in terms of gaps and limitations in building STFN.

Chapter 3: Research Methodology.

This chapter discusses the research methodology used to carry out the study. There are two methods proposed in building STFN for fuzzy data preparation, which is measurement error and standard deviation. The explanation for both approaches is presented in this chapter, including phases and steps that will be taken in this research to achieve the objectives.

Chapter 4: Simulation.

This chapter explains the simulation process of the proposed approaches for autoregressive prediction. There are two types of validation used in this chapter, which are parameter validation and Mean Square Error (MSE) validation. Furthermore, the tool used to produce the generated data is also presented in this chapter.

Chapter 5: Result and Analysis.

This chapter explains the implementation of the proposed approaches for real time series data. Eight datasets from two fields of study, which consist of the stock markets and Air Pollutant Index, were used for implementation. The performance evaluation was carried out based on MSE validation.

Chapter 6: Conclusion and Future Works.

The contributions of the proposed methods of building STFN for fuzzy data preparation of the time series analysis are summarised, and the recommendations are given for further continuation of the work.

REFERENCES

1. K. N. Barker, “Data collection techniques: observation.,” *American Journal of Hospital Pharmacy*, vol. 37, no. September, pp. 1235–1243, 1980.
2. J. J. Hox and H. R. Boeije, “Data Collection, Primary vs. Secondary,” *Encyclopedia of Social Measurement*. pp. 593–599, 2005.
3. S. Fosso Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, “How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study,” *International Journal of Production Economics*, vol. 165, pp. 234–246, 2015.
4. M. Armbrust *et al.*, “Spark SQL: Relational Data Processing in Spark,” *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data(SIGMOD)*, pp. 1383–1394, 2015.
5. M. Kearse *et al.*, “Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data,” *Bioinformatics*, vol. 28, no. 12, pp. 1647–1649, 2012.
6. B. S. Xia and P. Gong, “Review of business intelligence through data analysis,” *Benchmarking*, vol. 21, no. 2, pp. 300–311, 2014.
7. B. Ravel and M. Newville, “ATHENA, ARTEMIS, HEPHAESTUS: Data analysis for X-ray absorption spectroscopy using IFEFFIT,” *Journal of Synchrotron Radiation*, vol. 12, no. 4, pp. 537–541, 2005.
8. L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry, “Affy - Analysis of Affymetrix GeneChip data at the probe level,” *Bioinformatics*, vol. 20, no. 3, pp. 307–315, 2004.
9. S. L. R. Ellison and A. Williams, *Quantifying Uncertainty in Analytical Measurement*, vol. 2nd. 2000.
10. G. Mahalakshmi, S. Sridevi, and S. Rajaram, “A survey on forecasting of time series data,” *2016 International Conference on Computing Technologies and Intelligent Data Engineering, ICCTIDE 2016*, 2016.
11. Z. Ismail, R. Efendi, and M. M. Deris, “Inter-Quartile Range Approach To Length–Interval Adjustment of Enrollment Data in Fuzzy Time Series Forecasting,” *International Journal of Computational Intelligence and*

Applications, vol. 12, no. 03, p. 1350016, 2013.

12. S. M. Chen, H. P. Chu, and T. W. Sheu, "TAIEX forecasting using fuzzy time series and automatically generated weights of multiple factors," *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, vol. 42, no. 6, pp. 1485–1495, 2012.
13. S. M. Chen and P. Y. Kao, "TAIEX forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines," *Information Sciences*, vol. 247, pp. 62–71, 2013.
14. L. Y. Wei, C. H. Cheng, and H. H. Wu, "A hybrid ANFIS based on n-period moving average model to forecast TAIEX stock," *Applied Soft Computing Journal*, vol. 19, pp. 86–92, 2014.
15. M. Y. Chen and B. T. Chen, "A hybrid fuzzy time series model based on granular computing for stock price forecasting," *Information Sciences*, vol. 294, pp. 227–241, 2015.
16. G. P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting," *Information Sciences*, vol. 177, no. 23, pp. 5329–5346, 2007.
17. G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
18. W. Eugster, *Eddy covariance: A practical guide to measurement and data analysis*, no. May 2014. Springer Science & Business Media, 2012.
19. H. W. Coleman and W. G. Steele, "Experimentation, Validation and Uncertainty Analysis for Engineers," *Journal of Chemical Information and Modeling*, 2018.
20. S. S. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, *Measurement Errors in Surveys*. John Wiley & Sons, 2011, 2004.
21. A. Ferrero and S. Salicone, "Modeling and processing measurement uncertainty within the theory of evidence: Mathematics of random-fuzzy variables," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 3, pp. 704–716, 2007.
22. A. Ferrero and S. Salicone, "An innovative approach to the determination of uncertainty in measurements based on fuzzy variables," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 4, pp. 1174–1181, 2003.
23. F. Chau, R. Deesomsak, and J. Wang, "Political uncertainty and stock market

volatility in the Middle East and North African (MENA) countries," *"Journal of International Financial Markets, Institutions & Money,"* no. February, 2016.

24. C. Simon, P. Weber, and M. Sallak, *Data Uncertainty and Important Measures*, vol. 3. 2018.
25. D. Escobari and M. Jafarinejad, "Investors' Uncertainty and Stock Market Risk," 2018.
26. N. Arbaiy, N. A. Samsudin, J. Watada, and P. C. Lin, "An Enhanced Possibilistic Programming Model with Fuzzy Random Confidence-Interval for Multi-Objective Problem," *Innovative Computing, Optimization and Its Applications*, no. January, pp. 217–235, 2018.
27. R. Efendi, N. Arbaiy, and M. M. Deris, "A new procedure in stock market forecasting based on fuzzy random auto-regression time series model," *Information Sciences*, vol. 441, pp. 113–132, 2018.
28. P. C. Chang, J. L. Wu, and J. J. Lin, "A Takagi–Sugeno fuzzy model combined with a support vector regression for stock trading forecasting," *Applied Soft Computing*, vol. 38, pp. 831–842, 2016.
29. L. Maciel, F. Gomide, and R. Ballini, "Evolving Fuzzy-GARCH Approach for Financial Volatility Modeling and Forecasting," *Computational Economics*, vol. 48, no. 3, pp. 379–398, 2016.
30. P. Singh, "An Efficient Method for Forecasting Using Fuzzy Time Series," in *Emerging Research on Applied Fuzzy Sets and Intuitionistic Fuzzy Matrices*, IGI Global, 2017, p. 18.
31. Z. Ismail, R. Efendi, and M. M. Deris, "Application of Fuzzy Time Series Approach in Electric Load Forecasting," *New Mathematics and Natural Computation*, vol. 11, no. 03, pp. 229–248, 2015.
32. R. Efendi and M. M. Deris, "Prediction of Malaysian-Indonesian Oil Production and Consumption Using Fuzzy Time Series Model," *Advances in Data Science and Adaptive Analysis*, vol. 9, no. 1, pp. 1–17, 2017.
33. Y. Leu, C. P. Lee, and Y. Z. Jou, "A distance-based fuzzy time series model for exchange rates forecasting," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8107–8114, 2009.
34. A. Nureize and J. Watada, "A fuzzy regression approach to a hierarchical evaluation model for oil palm fruit grading," *Fuzzy Optimization and Decision Making*, vol. 9, no. 1, pp. 105–122, 2010.

35. S. T. Steyn, Douw G., Rao, *Air Pollution Modeling and its Application XX*. 2010.
36. N. Middleton *et al.*, "A 10-year time-series analysis of respiratory and cardiovascular morbidity in Nicosia, Cyprus: The effect of short-term changes in air pollution and dust storms," *Environmental Health: A Global Access Science Source*, vol. 7, pp. 1–16, 2008.
37. L. S. Senthilkumar, "Triangular Approximation of fuzzy numbers - a new approach," vol. 113, no. 13, pp. 115–121, 2017.
38. X. Liu and Y. Chen, "A Systematic Approach to Optimizing h Value for Fuzzy Linear Regression with Symmetric Triangular Fuzzy Numbers," *Mathematical Problems in Engineering*, vol. 2013, 2013.
39. G. Mansingh, K.-M. Osei-Bryson, L. Rao, and M. McNaughton, "Data preparation: Art or science?," *International Conference on Data Science and Engineering, ICDSE 2016*, 2017.
40. S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, p. 2003, 2010.
41. R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," *Springer*, vol. 1, no. March 2014, pp. 1–26, 1999.
42. M. Magnani and D. Montesi, "Uncertainty in data integration: Current approaches and open problems," *CTIT workshop proceedings series*, pp. 1–15, 2007.
43. P. G. Moore, "Measuring Uncertainty," *Teaching Statistics*, vol. 2, no. 1, pp. 2–7, 1980.
44. K. S. Park, *Human reliability: Analysis, prediction, and prevention of human errors.*, 7th ed. Elsevier, 2014.
45. Kiran, "A Collective Assessment on Data Manipulation in Research Science," *International Journal of Applied Engineering Research*, vol. 14, no. 3, pp. 829–834, 2019.
46. R. Efendi, M. M. Deris, and Z. Ismail, "Implementation of Fuzzy Time Series in Forecasting of the Non-Stationary Data," *International Journal of Computational Intelligence and Applications*, vol. 15, no. 02, p. 1650009, 2016.
47. L. A. Zadeh, "Fuzzy Sets," *Inform Control*, no. 8, pp. 338–353, 1965.
48. N. Gerami Seresht and A. R. Fayek, "Computational method for fuzzy

arithmetic operations on triangular fuzzy numbers by extension principle," *International Journal of Approximate Reasoning*, vol. 106, pp. 172–193, 2019.

49. Y. Wang and W. Ran, "Comprehensive eutrophication assessment based on fuzzy matter element model and monte carlo-triangular fuzzy numbers approach," *International Journal of Environmental Research and Public Health*, vol. 16, no. 10, 2019.
50. J. Dhivya and B. Sridevi, "A novel similarity measure between intuitionistic fuzzy sets based on the mid points of transformed triangular fuzzy numbers with applications to pattern recognition and medical diagnosis," *Applied Mathematics*, vol. 34, no. 2, pp. 229–252, 2019.
51. H. Karimi, M. Sadeghi-Dastaki, and M. Javan, "A fully fuzzy best–worst multi attribute decision making method with triangular fuzzy number: A case study of maintenance assessment in the hospitals," *Applied Soft Computing Journal*, no. xxxx, p. 105882, 2019.
52. C. Li, "A fuzzy multi-objective linear programming with interval-typed triangular fuzzy numbers," *Open Mathematics*, vol. 17, no. 1, pp. 607–626, 2019.
53. W. S. W. Daud, N. Ahmad, and G. Malkawi, "A new solution of pair matrix equations with arbitrary triangular fuzzy numbers," *Turkish Journal of Mathematics*, vol. 43, no. 3, pp. 1195–1217, 2019.
54. J. Xu, J. Y. Dong, S. P. Wan, D. Y. Yang, and Y. F. Zeng, "A Heterogeneous Multiattribute Group Decision-Making Method Based on Intuitionistic Triangular Fuzzy Information," *Complexity*, vol. 2019, no. 1, 2019.
55. N. Arbaiy and J. Watada, "Linear fractional programming for fuzzy random based possibilistic programming problem," *International Journal of Simulation: Systems, Science and Technology*, vol. 15, no. 1, pp. 26–32, 2014.
56. J. J. Carbajal-Hernández, L. P. Sánchez-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "Assessment and prediction of air quality using fuzzy logic and autoregressive models," *Atmospheric Environment*, vol. 60, pp. 37–50, 2012.
57. N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, "Time-series Extreme Event Forecasting with Neural Networks at Uber," *International Conference on Machine Learning*, vol. 34, pp. 1–5, 2017.
58. R. Efendi, N. Arbaiy, and M. M. Deris, "Indonesian-Malaysian Stock Market

Models Using Fuzzy Random Time Series," pp. 18–19, 2017.

59. F. Mirzaei Talarposhti, H. Javedani Sadaei, R. Enayatifar, F. Gadelha Guimarães, M. Mahmud, and T. Eslami, "Stock market forecasting by using a hybrid model of exponential fuzzy time series," *International Journal of Approximate Reasoning*, vol. 70, pp. 79–98, 2016.
60. H. Yan and Z. Zou, "Application of a Hybrid ARIMA and Neural Network Model to Water Quality Time Series Forecasting," *Journal of Convergence Information Technology (JCIT)*, vol. 8, no. 2, pp. 1–12, 2013.
61. A. Fretheim and O. Tomic, "Statistical process control and interrupted time series: A golden opportunity for impact evaluation in quality improvement," *BMJ Quality and Safety*, vol. 24, no. 12, pp. 748–752, 2015.
62. R. B. Penfold and F. Zhang, "Use of interrupted time series analysis in evaluating health care quality improvements," *Academic Pediatrics*, vol. 13, no. 6 SUPPL., pp. S38–S44, 2013.
63. K. Zhang, R. Gençay, and M. Ege Yazgan, "Application of wavelet decomposition in time-series forecasting," *Economics Letters*, vol. 158, pp. 41–46, 2017.
64. I. Ackah and F. Adu, "Modelling gasoline demand in Ghana: A structural time series approach," *International Journal of Energy Economics and Policy*, vol. 4, no. 1, pp. 76–82, 2014.
65. E. D. Landoh *et al.*, "Morbidity and mortality due to malaria in Est Mono district, Togo, from 2005 to 2010: A times series analysis," *Malaria Journal*, vol. 11, pp. 1–7, 2012.
66. S. Sanei and H. Hassani, "Singular Spectrum Analysis," *Singular Spectrum Analysis of Biomedical Signals*, no. 18354, pp. 17–50, 2016.
67. Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - Part I," *Fuzzy Sets and Systems*, vol. 54, no. 1, pp. 1–9, 1993.
68. hyi-M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets and Systems*, vol. 81, no. 3, pp. 311–319, 1996.
69. M. Geurts, G. E. P. Box, and G. M. Jenkins, "Time Series Analysis: Forecasting and Control," *Journal of Marketing Research*, vol. 14, no. 2, p. 269, 2006.
70. F. Tseng, G. Tzeng, H. Yu, and B. J. C. Yuan, "Fuzzy ARIMA model for forecasting the foreign exchange market," *Fuzzy sets and systems*, vol. 118, no. 1, pp. 9–19, 2001.

71. J. S. Clark and O. N. Bjornstad, "Population time series: process variability, observation errors, missing values, lags, and hidden states," *Ecology*, vol. 85, no. 11, pp. 3140–3150, 2013.

72. R. J. Hyndman and G. Athanasopoulos, *Forecasting : Principles and Practice*. OTexts, 2018.

73. G. Uthra, K. Thangavelu, and B. Amutha, "An Approach of Solving Fuzzy Assignment Problem using Symmetric Triangular Fuzzy Number," vol. 113, no. 7, pp. 16–24, 2017.

74. P. Lavrakas, *Encyclopedia of Survey Research Methods*. Sage Publications, 2008.

75. H. Zhai, "Linear and Non-linear Regression : Application to Competitor ' s Gasoline Volume Estimation," pp. 1–27, 2015.

76. S. H. Cheng, S. M. Chen, and W. S. Jian, "A Novel Fuzzy Time Series Forecasting Method Based on Fuzzy Logical Relationships and Similarity Measures," *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*, pp. 2250–2254, 2016.

77. D. Pyle, *Data Preparation for Data Mining*. Morgan Kaufman Publisher, 1999.

78. J. E. Hanke and D. W. Wichern, *Business Forecasting*. 2009.

79. J. Buonaccorsi, "Measurement Error: Models, Methods and Applications," *Chapman and Hall/CRC*.

80. J. S. Yao, J. S. Su, and T. S. Shih, "Fuzzy system reliability analysis using triangular fuzzy numbers based on statistical data," *Journal of Information Science and Engineering*, vol. 24, no. 5, pp. 1521–1535, 2008.

81. A. Gani, Nagoor and S. N. M. Assarudeen, "A new operation on triangular fuzzy number for solving fuzzy linear programming problem," *Applied Mathematical Sciences, ,* vol. 6 no. 11, no. September 2014, p. 525 – 532 A, 2012.

82. J. Wang, D. Ding, O. Liu, and M. Li, "A synthetic method for knowledge management performance evaluation based on triangular fuzzy number and group support systems," *Applied Soft Computing Journal*, vol. 39, pp. 11–20, 2016.

83. E. Atify, C. Daoui, and A. Boumezzough, "Using A Fuzzy Number Error Correction Approach to Improve Algorithms in Blind Identification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 3, no.

2, p. 410, 2016.

84. R. E. Shannon, *Systems Simulation: the Art and Science*. Prentice Hall, 1975.
85. R. E. Shannon, "Introduction To The Art And Science Of Simulation," *Proceedings of the 1998 Winter Simulation Conference*, vol. 1, pp. 7–14, 1998.
86. R. Efendi, Z. Ismail, and M. M. Deris, "Improved Weight Fuzzy Time Series As Used in the Exchange Rates Forecasting of Us Dollar To Ringgit Malaysia," *International Journal of Computational Intelligence and Applications*, vol. 12, no. 01, p. 1350005, 2013.
87. R. Efendi, Z. Ismail, and M. M. Deris, "A new linguistic out-sample approach of fuzzy time series for daily forecasting of Malaysian electricity load demand," *Applied Soft Computing Journal*, vol. 28, pp. 422–430, 2015.
88. D. J. Rumsey, *Statistics For Dummies , 2-eBook Bundle Statistics For Dummies*. 2011.
89. J.-B. du Prel, G. Hommel, B. Röhrig, and M. Blettner, "Confidence Interval or P-Value?," *Deutsches Aerzteblatt Online*, vol. 106, no. 19, pp. 335–339, 2009.
90. J. Watada, S. Wang, and W. Pedrycz, "Building confidence-interval-based fuzzy random regression models," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 6, pp. 1273–1283, 2009.
91. R. Efendi, N. Arbaiy, and M. M. Deris, "Estimation of Confidence-Interval for Yearly Electricity Load Consumption Based on Fuzzy Random Auto-Regression Model," *Computational Intelligence in Information Systems*, vol. 532, 2017.
92. I. Chinelo, *Fundamentals of Research Methodology and Statistics*. New Age International, 2016.
93. C. R. Kothari, *Research Methodology: Methods & Techniques*. New Age International, 2004.
94. P. Zannetti, *Air Pollution Modeling - Theories, Computational Methods and Available Software*. Springer Science & Business Media, 2013.
95. N. Funabashi *et al.*, "Recommended acquisition-parameters in achieving successful evaluation of coronary lumen patency surrounded by XIENCE of diameters < 3.0 mm in 1st generation 320-slice CT. XIENCE Phantom Study Part 1," *International Journal of Cardiology*, vol. 202, pp. 537–540, 2016.
96. M. S. Bergin, G. S. Noblet, K. Petrini, J. R. Dhieux, J. B. Milford, and R. A. Harley, "Formal uncertainty analysis of a Lagrangian photochemical air

pollution model," *Environmental Science and Technology*, vol. 33, no. 7, pp. 1116–1126, 1999.

97. S. R. Hanna, "Air quality model evaluation and uncertainty," *Journal of the Air Pollution Control Association*, vol. 38, no. 4, pp. 406–412, 1988.
98. R. C. Nethery and F. Dominici, "Estimating pollution-attributable mortality at the regional and global scales: Challenges in uncertainty estimation and causal inference," *European Heart Journal*, vol. 40, no. 20, pp. 1597–1599, 2019.
99. S. Gul, M. T. Khan, D. I. Khan, and S. U. Rehman, "Stock Market Reaction to Political Events (Evidence from Pakistan)," *Journal of economics and sustainable development*, vol. 4, no. 1, pp. 165–174, 2013.
100. R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news," *ACM Transactions on Information Systems*, vol. 27, no. 2, pp. 1–19, 2009.
101. Rajendran Sugumar Alwar Rengarajan Chinnappan Jayakumar, "A Technique To Stock Market Prediction Using Fuzzy Clustering," *Computing and Informatics*, vol. 33, no. 5, pp. 992–1024, 2014.
102. S. A. Basher, A. A. Haug, and P. Sadorsky, "The impact of economic policy uncertainty and commodity prices on CARB country stock market volatility," *Munich Personal RePEc*, no. 96577, 2019.
103. Y. K. Liu and B. Liu, "Fuzzy random variables: A scalar expected value operator," *Fuzzy Optimization and Decision Making*, vol. 2, no. 2, pp. 143–160, 2003.
104. K. L. Judd, "The law of large numbers with a continuum of IID random variables," *Journal of Economic Theory*, vol. 35, no. 1, pp. 19–25, 1985.
105. N. Etemadi, "An elementary proof of the strong law of large numbers," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 55, no. 1, pp. 119–122, 1981.